

Digital Imaging Standards and Best Practices
Hoosier History Digitization Grants
SHRAB.IN.gov

This document provides information on the application of published standards and best practices for digital imaging to determine specifications for individual projects. It includes general principles based on best practices and minimum scanning guidelines for various material types.

Indiana SHRAB-funded projects must adhere to these guidelines in order to receive funding.

The Indiana Digital Archives facilitates the uploading of digital images of scanned records or searchable indexes, which lead a patron to further information. Available images are scanned as TIFFs. More information is provided below.

- GENERAL PRINCIPLES -

Capture once, use many times

Digitization is expensive, time-consuming, and requires extensive handling of original materials. Any digitization project should thus focus on creating high-quality master images from which many derivative images can be created for specific uses (e.g., web delivery). The master image should capture all "important" information from the original material, which should be explicitly defined for each digitization project. The master image should also be flexible enough to allow derivatives to be created meeting a wide variety of current and future needs. Therefore, no image processing (such as sharpening) should be done to the master file. The best practices described in this document are designed to achieve this goal of flexibility.

Create a faithful reproduction of the original

Access images displayed via the Indiana Digital Archives should look as close as possible to the original material from which they were derived. Many cultural heritage materials are aged and may show signs of damage from years of wear and tear. For example, black and white photographs can yellow over time, or there may be breakage around the edge of the pages of a diary from the 1800's. It is tempting to eliminate those signs of aging by digitally erasing a tear in the paper, cropping items so as to hide damaged areas, or color-correcting images to get rid of evidence of sun damage or discoloration. It is important not to take these measures, as these "imperfections" are often of interest to artists, historians, archaeologists, archivists, and researchers for a variety of reasons. There are some red, orange, and fluorescent hues that will not look exactly like the original no matter what digital imaging software is used to achieve a color balance, but digital scanning technicians should do their best to ensure that digital surrogates are faithful to the look and condition of the original material.

Scan from earliest generation practical

As copies are made of analog materials, each generation loses some detail. From a photographic negative to a print to a copy negative, from a book to microfilm, there is generational loss of information. To capture the most information in a scanned image, always use the earliest generation of the original material that is practical to use. In general, scanning

from negatives rather than prints and scanning from original printed material rather than microfilm or photocopies is preferable. However, there are cases where practical considerations dictate using a second- or third-generation original as the source of a scanned image. A set of cracked or broken glass-plate negatives might benefit from professional printing, then scanning the prints. A large series of bound volumes that have been microfilmed would be considerably cheaper to scan from microfilm rather than to unbind the volumes for scanning or to invest in expensive face-up scanning equipment. In these cases, a determination must be made if images created from later-generation originals can still meet the flexibility goals of master images for the project.

Technical issues

When setting technical specifications for digitization projects, higher is not always better. There is no advantage to scanning at a resolution higher than what is needed to capture the amount of detail on the original. In fact, there is a large disadvantage to this practice in that this excess resolution adds file size without adding detail to the digital image. The guidelines in this document are designed to help determine appropriate specifications and ensure files are as large as they need to be, but no larger.

A digitization program should employ some sort of color management solution to ensure scanners, monitors, and printers all represent image color accurately. Using “canned” International Color Consortium (ICC) profiles for each imaging device is a low-cost, somewhat effective mechanism, while using professional profiling software is a much more accurate but higher-cost solution.

Using digitization equipment appropriate to the materials being scanned is essential to an effective digitization project. Unfortunately, there are no one-size-fits-all digitization equipment solutions. For example, flatbed scanners are useful for unbound textual materials and photographic prints, while transparencies and negatives are more effectively imaged with dedicated film scanners. Never use a scanner at a resolution setting above its listed optical resolution (known as an interpolated resolution).

Quality control

A structured quality control program is essential to a good digitization project. An effective program might combine automated checking of objective criteria such as image resolution, file size, dimensions, and bit depth for all images with manual checking of subjective criteria such as color fidelity on a subset of scanned images.

Technical metadata

Keep a record of the type of scanner and technical metadata used for scanning. Many scanners will automatically save this information to the data file, but not all do. Even though the technical metadata will not be required by the Indiana Digital Archives, work from the “scan once, use many times” principle.

- MINIMUM SCANNING GUIDELINES -

The Indiana Digital Archives requires a basic **300 dpi TIFF** master file. Files **should not** be saved in alternate formats such as bitmap (BMP), JPEG, DWTIFF, RAW, or ARW.

File sizes should be kept at minimum; the maximum file size allowed is 2GB, although this would be an extremely large image. Files should never be saved **with a bitmap or bitonal compression even in TIFF format**. Bitmaps or bitonal compressed files offer the patron little more than poor photocopy quality.

Again, considering the “scan once” principle, if a 600 dpi TIFF or higher is more desirable for additional long-term preservation projects, reduce the file to a 300 dpi TIFF before submission to the Indiana SHRAB.

To reduce the unnecessary use of storage space, minimize “dead” or blank space around a scanned object, and crop images as close to the original image as possible. Test the image files by opening them in Microsoft Windows Picture and Fax Viewer. If they will open here, they should open in the Digital Archives.

Once uploaded to the Digital Archives, the master TIFF files will be stored in their original form within the system, but JPEG copies will be automatically generated for web presentation. Therefore, it is **not** necessary to create additional smaller files for public use or as thumbnails during the scanning process.

Indexing and Metadata

In order to make records searchable by patrons, each set of data must be accompanied by information about the content, known as *metadata*. This information should include the names of the people involved, record dates, etc. Indiana State Archives staff can accept metadata in either Access tables or Excel spreadsheets for conversion and upload to the Indiana Digital Archives.

Before developing the indexing fields, please visit the Indiana Digital Archives (www.digitalarchives.in.gov) to see examples of how records are displayed and the type of metadata that should be recorded.

The Digital Archives is arranged by Record Series and, within each series, by Record Titles. Currently the State Archives uses the following four series:

- **Institution Records**
- **Military Records**
- **Misc Historical Records**
- **Naturalization Records**

Additional Record Series available, but not used, include: Audio Records, Birth Records, Census Records, *Death Records**, Marriage Records, Minute Records, Ordinance Records, Resolution Records, Oath of Office Records, Photographs, Real Property Records Cards. At present, most county records fall under Miscellaneous Historical Records or Military Records

*Death Records currently on the Indiana Digital Archives consist of the Social Security Index, which is imported by Washington State through their system.

Fields

When indexing your records, the metadata fields you use will vary depending on the Record Series. It's best to include as much information in your index as possible to maximize its value to researchers, but the following fields are those required or recommended for every record series as a minimum.

1) LAST NAME / FIRST NAME

At this time all record series on the Indiana Digital Archives, with the exception of PHOTOGRAPHS, can only be searched by LAST NAME and/or FIRST NAME. Therefore, records **must** be indexed by these two fields if they are to be included.

When indexing your records, LAST name and FIRST name must be entered as separate fields in your spreadsheet/database. MIDDLE names may be combined with FIRST names (see the example below); alternatively, they can be added as separate fields, although these will not be searchable.

2) COUNTY

Records are not typically searchable by county. In order to be able to search by county within a record series, the data for each county needs to be split into separate tables before submission to the Indiana SHRAB

As an example of this, look at the "Naturalizations" record series. The data for each County is added under a separate TITLE. This allows a patron to search for a person within a specific county after opening that TITLE. However, all naturalization records can be searched at once by simply beginning the search from the main Naturalization page.

It is important that the county, if known, is included regardless of whether the data is separated by county, as this information will still be viewable to researchers on the results page. Having the county of origin is often very helpful to patrons.

3) NOTES

A notes field is available under each Record Series and Title. This can be very useful for providing further description about a record.

4) IMAGE

Image numbers within a record title must be unique to that Record Title.

001.tiff

002.tiff

If a record has multiple images associated with it—for instance, several pages in a book—it should be scanned as a multi-page TIFF.

Some equipment or software does not allow for creating multipage formats. A free, user friendly, open source software program *Irfanview* is available online at www.irfanview.com. This allows combining TIFF files to create a multi-page format.

Each set of images should relate specifically to a record of metadata. Enter the number of the new multi-page image in a column titled “IMAGE.”

5) DATE

The date field is only required under the Record Series “INSTITUTIONS.” However, researchers will benefit by having at least an approximate date for every record.

If no date is available insert the code “nd” (without the quotation marks).

Keep in mind that Microsoft Excel does not work well with true date fields in the 19th century. It is best to use a text field when entering dates in the 1800’s. 20th century dates in an Excel format will work in the Digital Archives.

6) REFERENCE NUMBER

All Record Series must have this field. Please contact Archives staff for assistance in developing appropriate Reference Numbers, and to ensure numbers are not already in use. 317-591-5222 or arc@iara.IN.gov